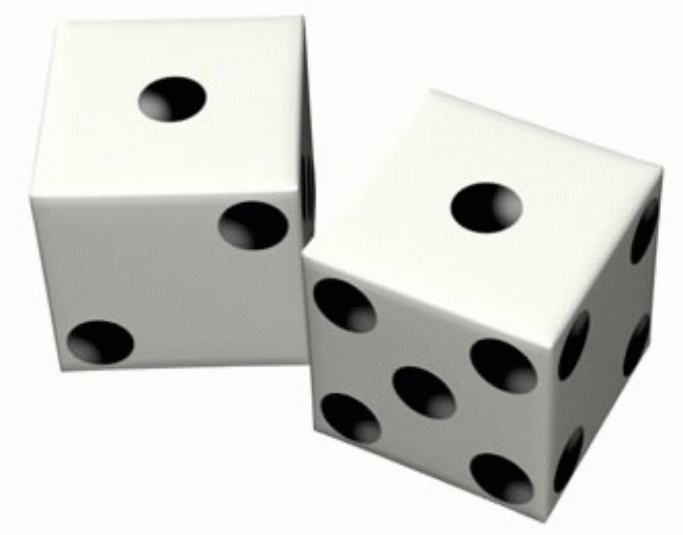
ASTR 600

STATISTICAL ANALYSIS IN PHYSICS AND ASTRONOMY

Spring Semester 2015 Room HBH 254; MWF 1:00-2:00



Instructor: Dr. Patrick Hartigan, Hermann Brown Rm. 350 , Phone: X2245

Office Hrs: After class and by appt.

Texts:

- *Modern Statistical Methods for Astronomy (with R applications)* by E. Feigelson and G. Babu
  This will be our main text, and students should own a copy.
- *REA's Statistics Problem Solver* A really nice set of worked-out examples for various statistical tests.

- Research Journals in Physics and Astronomy
  http://adsabs.harvard.edu/abstract/_service.html is a good source for the astronomy journals

- There are many statistics reference books listed at the end of each chapter in the Feigelson and Babu book

Grading: Based on short and long class presentations (50%), homework creation and completion (30%), and participation (20%)

Prereqs: There are no statistical prerequisites. We will cover material as we go.

OUTLINE OF COURSE

Every discipline of science has data, and a researcher generally wants to address a scientific question with the data. The path from the data to the conclusion usually involves some sort of statistical analysis. Even if the problem is not inherently a statistical one, when writing up the results one must always address the question of how certain the conclusions are given that there

are uncertainties in the measurements. Often one looks for correlations between different variables, or tries to infer what the underlying structural variables are that define the system, and how they relate to one another. The statistical literature is vast and different disciplines apply different techniques that are suitable to their data. But in most cases the techniques are quite broad, and in principle may be used in many different contexts.

In this class we will identify the statistical methods commonly used by research papers in astrophysics, and the students will chose papers to study that use these methods. By learning how statistics is used in different areas, we will broaden our toolset for use in our own studies. We will have short discussions on many such papers to get an overview of the usage, and then more in-depth discussions that will be led by students on their favorite methods. To solidify this knowledge, students will sometimes construct problems that will be distributed to the remainder of the class, and after a period of time we will discuss the solutions.

The topics will be guided by the professor, but the papers are chosen by the students. Some examples may include hypothesis testing and confidence intervals, extracting periods and signals from time-series data, using principal component analysis, maximum likelihood methods, finding groups and clusters in multidimensional data, maximum entropy reconstructions, non-parametric ranks, optimized profile extractions, Bayesian analysis with priors and so on. Most statistical methods used in astronomy will fall under one of the broad topics in this course, but if you have seen a technique and wondered how it works, we can substitute it for one of the other topics if there is interest. find a paper that uses it and we will study it.

In class we will, (*) find a broad range of papers chosen by the students that use a particular method, (*) study the mathematics of the method, (*) go in-depth into two papers chosen by the students, and (*) work through some problems to understand how the method works in practice. The students will pick the papers, and will be expected to lead the class through an overview of the science objectives. The main focus, however, will be on the method, which we will then all work together to try to understand. The problems may apply the method to some other case, perhaps with contrived data, so we are all sure we could use it if the need arose in another context. There will be a few lectures at the beginning and then interspersed as needed throughout the semester to provide some mathematical structure, and we will follow the overall format of the textbook. But the choice of specific applications is student-driven. The papers need not necessarily be astronomical, though that will likely be the focus of much of our studies.

Work Load, Absence and Late Policies:

- Every 1.5 weeks: 10-minute presentation to class on an article that uses the current technique we are studying
- Every 1.5 weeks: Complete a homework problem devised in part by a classmate
- Roughly every six weeks: A 40-minute presentation to the class that goes through the details of a paper, re-analyzing data or some subset of it if possible.
- Roughly every six weeks: Work with prof to create a homework problem with the solution for the technique we are using
- Absences: The class relies heavily on presentations and on homework problems that the students create. As such, participation grades rely heavily upon the presentations and homework being ready when needed. This is especially true for the long presentations. Makeups may not be possible given the schedule. Homeworks must be turned in on time to receive credit, because we discuss them on the date they are due.

Outcomes/Assessments

(This section is now required by the Rice Administration for all Syllabi)

Students completing this class should be able to do the following:

- Understand the fundamentals of statistics, including probability distributions, means, variances, the Central Limit Theorem, hypothesis testing, error propagation, Bayesian analysis, jacknife, and bootstrap
- Understand modern statistical methods that relate to curve fitting, hypothesis testing, cluster analysis, principal component analysis, and time-series data
- Assess the veracity of statistical conclusions drawn from any observations or data in the natural sciences, social sciences, or humanities (subject areas chosen by the students)
- Evaluate whether or not the best statistical methods were used to analyze data in papers published in recent refereed journals
- Apply the statistical concepts covered in class to their own sets of data using, in part, the statistics software package R
- Create homework problems that relate to one of the mathematical techniques studied in class
- Improve speaking and presentation skills while leading class discussions
- Learn to summarize complex papers in a fixed amount of time

SPRING 2015 Schedule

DATE	Topic	Class Type	Subject and Items Due
W Jan 21	Mathematical Foundation	Lecture	Probability Distributions Chapter 2
F Jan 23	"	"	"
M Jan 26	"	"	Confidence Intervals Chapters 3 and 4
W Jan 28	"	"	"
F Jan 30	"	"	HMWK#1 handout **[pmh]**
M Feb 2	Nonparametrics	Short presentations **[Everyone]**	Paper discussions
W Feb 4	"	Lecture	Chapter 5
F Feb 6	"	"	"
M Feb 9	"	"	*HMWK#1 due* **[Everyone]**; discussion
W Feb 11	"	Long paper **[Willie]**	Paper discussion + HMWK#2 handout **[Amanda]**
F Feb 13	Regression	Short presentations **[Everyone]**	Paper discussions
M Feb 16	"	Lecture	Chapter 7
W Feb 18	"	"	"
F Feb 20	"	"	*HMWK #2 due* **[All but Amanda]** and discussion

M Feb 23	"	Long paper **[Andy]**	Paper discussion + HMWK#3 handout **[Josh]**
W Feb 25	Multivariate, PCA	Short presentations **[Everyone]**	Paper discussions
F Feb 27	"	Lecture	Chapter 8
M Mar 2 - F Mar 6	Midterm RECESS	[NO CLASS]	---
M Mar 9	"	"	"
W Mar 11 - F Mar 13	pmh and Andy on travel	[NO CLASS]	---
M Mar 16	"	"	*HMWK #3 due* **[All but Josh]** and discussion
W Mar 18	Multivariate PCA	Long paper **[Amanda]**	Paper discussion + HMWK#4 handout **[Willie]**
F Mar 20	Time-Series Analysis	Short presentations **[Everyone]**	Paper discussions
M Mar 23	"	Lecture	Chapter 11
W Mar 25	"	"	"
F Mar 27	"	"	*HMWK #4 due* **[All but Willie]** and discussio
M Mar 30	"	Long paper **[Josh]**	Paper discussion + HMWK#5 handout **[Andy]**
W Apr 1	Clustering	Short presentations **[Everyone]**	Paper discussions
F Apr 3	Easter RECESS	[NO CLASS]	---
M Apr 6	"	Lecture	Chapter 9
W Apr 8	"	"	"
F Apr 10	"	"	*HMWK #5 due* **[All but Andy]** and discussion
M Apr 13	"	Long paper **[Willie]**	Paper discussion + HMWK#6 handout **[Amanda]**
W Apr 15	Truncated Data	Short presentations **[Everyone]**	Paper discussions
F Apr 17	"	Lecture	Chapter 10
M Apr 19	"	"	"
W Apr 22	"	"	*HMWK #6 due* **[All but Amanda]** and discussion
F Apr 24	"	Long paper **[Andy]**	Paper discussion + HMWK#7 handout **[Josh]**
M May 1	"		*HMWK #7 due* **[All but Josh]**

Honor Code: A general description of the honor code is avilable on-line. Students should turn in their own work and analysis on the homework sets, but may discuss the general nature of the problems with one-another.

Disability Accommodation: If you have a documented disability that will impact your work in this class, please contact me to discuss your needs. Additionally, you will need to register with the Disability Support Services Office in the Ley Student Center.

Short summary of techniques